DOCUMENT RESUME

ED 459 195 TM 033 483

TITLE	Addendum to Improving Classroom Assessment: A Toolkit for Professional Developers.
INSTITUTION	Northwest Regional Educational Lab., Portland, OR.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	2000-09-00
NOTE	56p.; For the Toolkit itself, see ED 423 253. For the full text of the Toolkit, see http://www.nwrel.org/eval/Toolkit98.
AVAILABLE FROM	Northwest Regional Educational Laboratory, 101 S.W. Main Street, Suite 500, Portland, OR 97204. Fax: 503-275-0458; Web site: http://www.nwrel.org.
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Educational Assessment; Elementary Secondary Education; *Grading; *Performance Based Assessment; Standards; Student Evaluation; *Test Construction; Test Use
IDENTIFIERS	*Alternative Assessment

ABSTRACT

This Addendum was created with the expectation that it would enhance the understanding of the conceptual framework of "Improving Classroom Assessment: A Toolkit for Professional Developers," link its text and activities to other assessment aids, and help trainers design high-quality workshops. The first section of the addendum contains an overview of the "Toolkit" that can be background reading for trainers or material to help others understand the structure of the Toolkit. Section 2, "Extending the Concepts Built by the 'Toolkit'", provides a set of cross-referencing tools that assist in finding the right tools and strategies to accomplish objectives for quality assessment. Section 3, "Tools To Help You Plan Your 'Toolkit' Training," contains materials to assist in designing training activities. Section 4 contains replacement pages for chapter 1 of the Toolkit. (SLD)

Addendum to Improving Classroom Assessment: A Toolkit for Professional Developers

September 2000

Regional Educational Laboratories

The REL Assessment Laboratory Network Program (Assessment Program LNP)

Nationwide, the Regional Educational Laboratories (REL) collaborate to form the REL Assessment Program LNP (Laboratory Network Program), dedicated to improving student learning on a national scale by providing access to resources, technical assistance, and professional development in the area of classroom assessment.

Products and publications developed by the REL Assessment Program LNP include:

- ☐ *Improving Classroom Assessment: A Toolkit for Professional Developers* is 1,200+ power-packed pages including text, 37 training activities instructions, teacher-friendly readings about classroom assessment issues, and sample assessments representing a variety of design options. All can be used to promote discussion and conduct assessment training. This document can be found online at www.nwrel.org/eval/*Toolkit*98.

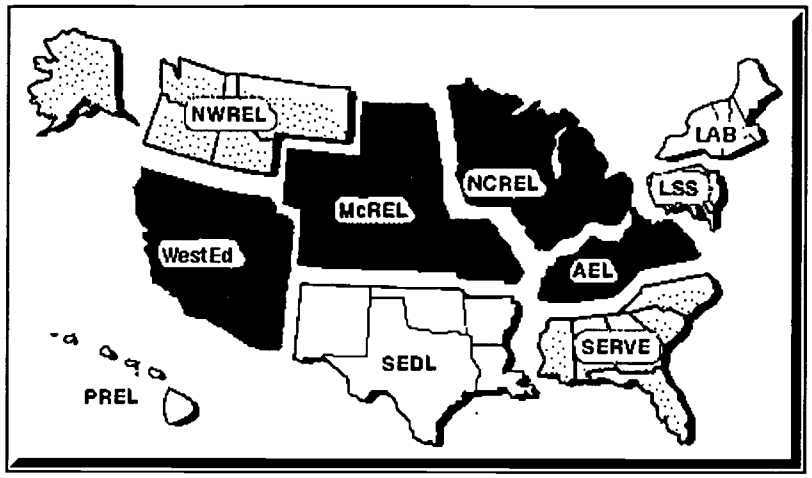
- ☐ *Making Assessment Work for Everyone: How to Build on Students' Strengths* is a self-study guide for teachers and administrators that provides information, resources, and activities for selecting, adapting, and developing assessments to promote excellence in each student while honoring all students' culture(s), experiences, and ways of knowing and showing learning. (See online at www.sedl.org/pub/tl02.)

- ☐ *The Assessment Program LNP Web Site*, at www.wested.org/acwt/, houses information about existing and upcoming Assessment Program LNP products and events.

- ☐ *The Promising Practices in Assessment Database (PPAD)*, accessible through the Assessment Program LNP Web site, is a searchable online database containing a variety of high quality, assessment-focused materials developed by the 10 regional educational laboratories.

- ☐ *REL Assessment Software Database,* accessible through the Assessment Program LNP Web site, is a searchable online database providing detailed information on assessment software that focuses on applications for gradebooks, test generation, resource assistance, electronic portfolios, and monitoring of student progress.

For further information about any of these products or services, please contact your Assessment Program LNP core work group representative.



Arlene Mitchell
McREL
2550 S. Parker Road, Suite 500
Aurora, CO 80014
Dburger@mcrel.org
(303) 337-0990
(303) 752-6388 (FAX)

Chris Ferguson
SEDL
211 East Seventh Street
Austin, TX 78701-3281
Cferguso@sedl.org
(512) 476-6861
(800) 476-6861
(512) 476-2286 (FAX)

Melissa Ritter
NWREL
101 SW Main Street, Suite 500
Portland, OR 97204-3297
Ritterm@nwrel.org
(503) 275-9562
(503) 275-0450 (FAX)

Mahna Schwager, Coord.
WestEd
730 Harrison Street
San Francisco, CA 94107
mschwag@wested.org
(415) 615-3201
(415) 615-3200 (FAX)

Jane Hange
AEL
P.O. Box 1348
Charleston, WV 25325
hangej@ael.org
(304) 347-0411
(304) 347-0487 (FAX)

Jeong-Ran Kim
LSS—Temple University
9th Floor,
Ritter Annex (004-00)
Philadelphia, PA 19122
email: jkim3@vm.temple.edu
(215) 204-3006
(215) 204-5130 (FAX)

Eileen F. Ferrance
The LAB at Brown University
222 Richmond Street, Suite 300
Providence, RI 02903-4226
Eileen_ferrance@brown.edu
(401) 274-9548 Ext. 228
(401) 421-7650 (FAX)

Sharon Horn
OERI
555 New Jersey Avenue NW
Washington, DC 20208
(202) 219-2203
Sharon_Horn@ed.gov

Nancy McMunn
SERVE
915 Northridge Street, 2d Floor
Greensboro, NC 27403
Nmcmunn@SERVE.org
(336) 315-7400
(800) 755-3277
(336) 315-7457 (FAX)

Arie van der Ploeg
NCREL
1900 Spring Road, Suite 300
Oak Brook, IL 60521
arie@ncrel.org
(630) 571-4700
(630)-571-4716)

Donald L. Burger
Monica Mann
PREL
1099 Alakea Street, Suite 2500
Honolulu, HI 96813
Burgerd@prel.org
Mannm@prel.org
(808) 533-6000
(808) 533-7599 (FAX)

As assessment trainers themselves, Toolkit development team members frequently test and share ideas, strategies, and documents. They created this Addendum with the expectation that the materials included will enhance understanding of the conceptual framework of the Toolkit, link its text and activities to other assessment aids, and help trainers design high-quality workshops. Each of the Assessment Lab Network Program representatives contributed information and materials refined through use at various trainings. They hope you find these materials as useful as they have.

Table of Contents

I. *Toolkit* Overview

Why Examine Classroom Assessment Training for Teachers?

Even though teachers can spend as much as a third to a half of their professional time involved in assessment-related activities (Herman & Dorr-Bremme, 1982; Crooks, 1988; Stiggins & Conklin, 1991), study after study shows that K-12 teachers lack skill in assessing their students (Impara, Plake, & Fager, 1993; Plake, Impara, & Fager, 1993; Hills, 1991) and feel unprepared and uncomfortable in their own knowledge of assessment practices (Shafer, 1993; Wise, Lukin, & Roos, 1991; Zhang, 1997). Due to current emphasis on standards-based education, teachers will probably be expected to be even more assessment literate in the future than they are today.

As Joan Herman (1997) points out (somewhat paraphrased):

- Assessments are the final word in defining what we want students to know and be able to do. Regardless of what we say in our standards documents (or in our course outcomes), the assessments define what we really mean. So, assessments communicate the standards to which school systems, school, teachers and students aspire.
- These standards provide focus and direction for teaching and learning.
- Results from assessments support important insights on the nature, strengths, and weaknesses of student progress relative to the standards.
- Educators and students must use this feedback to understand and direct their attention to improving relevant aspects of student learning.

In addition to the standards-based education movement, teachers need to be assessment literate because assessment affects kids. The influence of assessment on instruction, teachers, and students can be positive or negative. Just consider some of the negative consequences from past over-reliance on standardized multiple-choice tests. "Under pressure to help students do well on such tests, teachers and administrators have tended to focus their efforts on test content, to mimic the tests' multiple-choice formats in the classroom, and to devote more and more time to preparing students to do well on the tests. The net effect was a narrowing of the curriculum to the basic skills assessed and a neglect of complex thinking skills and other subject areas which were not assessed (Corbett & Wilson, 1991; Dorr-Bremme & Herman, 1986; Kellaghan & Madaus, 1991; Herman, 1997).

In a more personal look at the impact of classroom assessment on students, consider: "What we choose to evaluate and how we choose to evaluate delivers powerful messages to students about those things we value. Students view their learning and their sense of worth through the lens we help them construct unless they cannot bear to look through it (Stayter & Johnston, 1990).

These facts pose a significant challenge to those of us helping teachers acquire the student assessment competencies they need. How many of us work in a state where teachers are not even required to have assessment coursework to be certified? And where such coursework is required to be certified, is it a seat time requirement or a *competency* requirement? And when it's *competency*, competency in what—what do teachers really need to know and be able to do and at

what level? A recent study showed that, where preservice coursework is required or offered, there is a certain amount of feeling that it doesn't cover what teachers will really need to know and be able to do when they are teachers, or that the courses are not taught by those most familiar with assessment issues and developments (Fager, Plake, & Impara, 1997).

Given the importance of assessing well, we must all continue to think about and discuss what teachers need to know and be able to do with respect to classroom assessment, the best ways to assist them to learn it, and how we'll know when they're competent. This session is dedicated to a continuing exploration of these topics.

References:

Corbett, H.D. & Wilson, B.L. (1991). Testing, reform, and rebellion. Norwood, NJ: Ablex Publishing.

Crooks, T.J. (1988). The impact of classroom evaluation on students. *Review of Educational Research,* 58(4), 438-481.

Dorr-Bremme, D., & Herman, J. (1986). Assessing student achievement: A profile of classroom practices. *CSE Monograph Series in Evaluation No. 11.* Los Angeles: University of California, CRESST.

Fager, J.J., Plake, B.S., & Impara, J.C. (1997). *Examining teacher educators' knowledge of classroom assessment: A pilot study.* Paper presented at NCME national conference, Chicago, IL.

Herman, Joan (1997). *Assessing new assessments: How do they measure up?* CRESST, UCLA Graduate School of Education and Information Studies.

Herman, J. & Dorr-Bremme, D. (1982). *Assessing students: Teachers' routine practices and reasoning.* Paper presented at AERA, New York.

Hills, J.R. (1991). Apathy concerning grading and testing. *Phi Delta Kappa*, 72(7), 540-545.

Impara, J.C., Plake, B.S., & Fager, J.J. (1993). Teachers' assessment background and attitudes toward testing. *Theory into Practice*, 32(2), 113-117.

Kellaghan, T. & Madaus, G. (1991, November). National testing: Lessons for America from Europe. *Educational Leadership*, 49, 87-93.

Plake, B.S., Impara, J.C., & Fager, J.J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-12, 39.

Shafer, W.D. (1993). Assessment in teacher education. *Theory into Practice*, 32(2), 118-126.

Stayter, F., & Johnston, P. (1990). Evaluating the teaching and learning of literacy. T. Shanahan (Ed.), *Reading and writing together: New perspectives for the classroom*, Norwood, MA: Christopher-Gordon.

Stiggins, R. & Conklin, N. (1991). *In teachers' hands: Investigating the practice of classroom assessment.* Albany, NY: SUNY Press.

Wise, S.L., Lukin, L.E., & Roos, L.L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education*, 42(1), 37-42.

Zhang, Zhicheng (1997). *Assessment practices inventory: A multivariate analysis of teachers' perceived assessment competency.* Paper presented at NCME national conference, Chicago, IL.

Frequently Asked Questions about the *Toolkit*

1. **How can I get additional copies of the *Improving Classroom Assessment Toolkit*?**
   You can order additional copies of the *Toolkit* through the Regional Educational Laboratory (REL) that serves your area. Every *Toolkit* comes in two shrink-wrapped, three-hole-punched sections, along with a *CD*-ROM with masters for all handouts and overheads. The entire *Toolkit* is online at www.nwrel.org/eval/*toolkit*98/. Each chapter, training activity, and sample assessment can be individually downloaded and printed out or saved as an electronic file to print out for use.

2. **May I make copies of the *Toolkit* materials?**
   The *Toolkit* was developed with funding from the U.S. Department of Education, Office of Educational Research and Improvement. Permission is granted to all educators to copy portions for use in local professional development activities. When training others to use the *Toolkit*, the authors often make copies of all activities used in the training to share with participants.

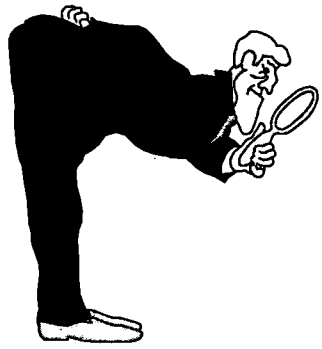
3. **How can I tailor the overheads to my training?**
   The *Toolkit* is designed to be flexible and adaptable. As you use various *Toolkit* materials, you may want to make changes so that your overheads fit the training that you have planned. Perhaps your school doesn't use the term "learning targets" but rather "instructional targets." Using terminology familiar to your audience is helpful in training. The disk included with a purchased *Toolkit* contains all of the overheads for the *Toolkit* formatted using PowerPoint. This will enable you to tailor the overheads to your needs. However, when a credit or citation appears, that line should remain on the page. For example, Activity 1.7 (PPAct1-7.ppt on the *CD*-ROM) includes an overhead adapted from the work of Rick Stiggins of the Assessment Training Institute, and the credit line should remain on the page to recognize his work even when minor changes are made. The *CD*-ROM will enable you to make changes; the downloadable files on NWREL's Web site are not modifiable.

4. **How do I make printouts using the REL *Improving Classroom Assessment Toolkit CD*?**
   To open the *CD*-ROM with the PowerPoint-formatted overheads, your computer must have Windows95 or higher. When you order your *Toolkit*, be sure to specify the right platform, Mac or PC, for your computer. You will want to open and read the "Read Me" file included on the disk before trying to open any of the PowerPoint files.

   If you have the PowerPoint software on your computer, simply open the application and access the *Toolkit CD*-ROM through PowerPoint.

   If you don't have the PowerPoint software on your computer, you can still use the *Toolkit* CD-ROM. A PowerPoint viewer is included on the disk. This viewer will allow you to view, project, and print the overheads. However, it will not allow you to manipulate the order or content of the overheads; and, it won't allow you to download files to a word processing program.

Toolkit Overview



- Text, professional development activities (37), sample assessments (48), articles on grading (13)
- Each activity comes complete with handouts, hard copies of overheads, and a facilitator's guide
- Mix-and-match sections
- Not meant as a complete textbook on assessment

Toolkit Content Shortie

Chapter 1: Standards-Based Assessment–Nurturing Learning	Chapter 3: Designing High-Quality Assessments
Big ideas: • What is assessment? • What is quality assessment? • What is the role of assessment in standards-based education?	*Big ideas:* • What are the characteristics of high- quality alternative assessments? • What are the options for designing performance tasks and criteria? • When should these options be used?
Chapter 2: Integrating Assessment with Instruction	**Chapter 4: Grading and Reporting–A Closer Look**
Big ideas: • What does it mean to integrate assessment and instruction? • How can performance assessments be used to promote student learning of the very skills also being assessed?	*Big ideas*: • Why do we grade? • What are current grading and reporting dilemmas and issues? • What are alternatives to grading to report student learning?

Appendix A: Alternative Assessment Sampler: 48 sample alternative assessments in various grade levels and subject areas. (Used in various training activities.)

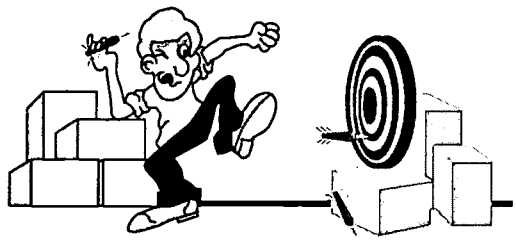
Appendix B: Student Work Samples: Seven sets of student work in various grade levels and subject areas. These are used in **Activity 2.1**—*Sorting Student Work.*

Appendix C: Articles on Grading: Thirteen articles on grading used in **Activities 4.1** and **4.2.**

Appendix D: Training Agenda Examples and Evaluation Forms. Sample ways that individuals have used *Toolkit* activities and sample forms to evaluate assessment training.

CHAPTER REVIEWS

Chapter 1
Standards-Based Assessment—Nurturing Learning



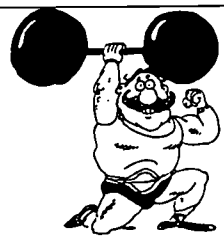
Chapter Goals

1. Establish background knowledge about assessment
2. Visit the notion of standards-based education and the role of assessment in the standards process
3. Increase awareness of the principles of good assessment
4. Lay the groundwork for good assessment as a tool for educational improvement
5. Agree on a common language of assessment terms

Readings

Activities

Chapter 2
Integrating Assessment With Instruction



Chapter Goals

1. Present different conceptions of what it means to integrate assessment and instruction
2. Discuss various ways that assessment can influence teachers and students
3. Assist the readers to build a vision of what they would like assessment to accomplish
4. Discuss the assessment design implications of various visions

Readings

Activities

Chapter 3
Designing High-Quality Assessments

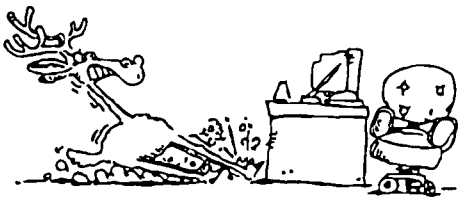
Chapter Goals

1. Illustrate design options for alternative assessments
2. Expand expertise in developing alternative assessments
3. Present guidelines for assessing the quality and appropriateness of alternative assessments for particular purposes and contexts
4. Provide experience in choosing high-quality assessments
5. Examine issues of equity, fairness, bias, and unintended consequences of assessment

Readings

Activities

Chapter 4
Grading and Reporting—A Closer Look



Chapter Goals

1. Explore issues with respect to grading and reporting
2. Reinforce the importance of aligning grading and reporting strategies with valued learning targets and instructional strategies
3. Increase knowledge of strategies for incorporating alternative assessment data in the overall determination of student grades
4. Reflect on options for broadening communication with parents, students, other teachers, and the community about the quality of student work

Readings

Activities

II. Extending the Concepts Built by the *Toolkit*

Cross-References that Will Prove Useful in Planning to Use the *Toolkit*

The following resource provides a grounded explanation for the conceptual framework for the *Toolkit*. Reading this document, available through ERIC, and using the cross-reference charts will assist you in understanding the rationale for the *Toolkit*.

ERIC NO: ED323186

TITLE: Standards for Teacher Competence in Educational Assessment of Students.

PUBLICATION DATE: 1990

ABSTRACT: The assessment competencies set forth in this monograph are knowledge and skills critical to a teacher's role as an educator. It is suggested that the seven standards described as essential for educational assessment of students be incorporated into future teacher training and certification programs. The standards require that teachers be skilled in the following competencies:(1) choosing assessment methods appropriate for instructional decisions; (2) developing assessment methods appropriate for instructional decisions; (3) administering, scoring, and interpreting the results of both externally produced and teacher-produced assessment methods; (4) using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement; (5) developing valid pupil grading procedures which use pupil assessments; (6) communicating assessment results to students, parents, other lay audiences, and other educators; and (7) recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information. (JD)

PAGE: 7; 1

Cross-Reference–Keys to Quality Assessment: *Toolkit*

The *Five Keys to Quality Assessment* represent an organizing principle for the *Toolkit*. If knowledge of the Keys is vital for teachers in creating assessments, then *Toolkit* activities can assist them to gain this knowledge. The chart below illustrates the relationship between *Toolkit* text and activities and the Five Keys.

Text	Key 1: What	Key 2: Why	Key 3: How	Key 4: How Much	Key 5: How Good
Chapter 1 Text	Contains an explicit discussion of all five Keys.				
Activity 1.1, Handout 1.1	Line 3 alludes to Key 1.	Lines 1, 2, 7, and 8 present view of purposes for assessment.	Lines 3, 4, and 6 allude to need for target-method match.	Line 4 alludes to sampling.	Lines 3 and 6 discuss several possible sources of bias and distortion.
Activity 1.2	Major focus.				
Activity 1.3	This is an introductory vocabulary activity.				
Activity 1.4		Assessment to help students learn.			Includes statements of quality.
Activity 1.5	Need for clear targets arises in discussion.	Need for a clear purpose arises in discussion.			Major focus.
Activity 1.6	Nature of targets arises in discussion.	Purposes arise in discussion.	Methods depend on what you assess and why you assess.		
Activity 1.7			Major focus.		
Activity 1.8	Be clearer than "level of proficiency" in choosing method.	Poses question if purpose determines preferred method.	Method has to match targets and purposes.	Need for multiple measures for total picture of student's ability.	There are flaws in every method proposed— one needs multiple measures.
Activity 1.9		Emphasizes need to have clear purposes to choose the best rubric type.	See Key 2.		
Activity 1.10	Major focus.		Major focus.		
Activity 1.11	Assessment standards in this activity state the Five Keys in slightly different words. For example, on p. 9, "multiple sources of assessment information should be used" is the same as Key 5.				
Activity 1.12, Handout 1	A4 restates the need for clear targets.	A1, A2, and A8 state the authors' main purposes for assessment.	A5 and A9 allude to appropriate assessment methods.	A5 also mentions sampling.	A3, A6, A7, and A8 all discuss possible sources of bias and distortion.
Chapter 2 Text		Assessment as a tool for learning and monitoring progress.			
Activity 2.1	Examines the nature of the targets.	Probes assessment purposes.	Rubric designs and various purposes.		

Text	Key 1: What	Key 2: Why	Key 3: How	Key 4: How Much	Key 5: How Good
Activity 2.2	See Activity 2.1	See Activity 2.1	See Activity 2.1		
Activity 2.3, Handout 1	Group 1 discussion.	Group 2 and 4 Purposes for assessment. Group 3 Understandable feedback for users and uses.			Group 5 Nature of quality, "authentic" tasks.
Activity 2.4		Types and purposes of assessments.			Criteria for good quality assessment tasks.
Activity 2.5		Performance criteria as a tool for communication and learning.			
Chapter 3 Text			Design depends on purposes, users and uses, and targets.		Quality considerations.
Activity 3.1			Same.		Same.
Activity 3.2			Same.		Same.
Activity 3.3			Same.		Same.
Activity 3.4			Same.		Same.
Activity 3.5			Same.		Same.
Activity 3.6	This activity requires participants to consider all five Keys in order to critique an assessment.				
Chapter 4 Text	Quality grading and reporting requires clear targets.	Formats, information, purposes, learning, and support.	Matching methods to targets.	Good sampling.	Minimization of bias and distortion.
Activity 4.1	Same.		Same.		
Activity 4.2		Same.			Same.
Activity 4.3	Same.	Same.	Same.	Same.	Same.
Activity 4.4	Same.	Same.			
Activity 4.5	Same.	Same.			
Activity 4.6		Purposes for assessment & grading.			Same.
Activity 4.7	Dilemma – what students are to know and be able to do.	Users and uses.			
Activity 4.8, Handout 1	Principle 4.	Principles 1 and 8.		Principle 2.	Principles 1, 3, 5, 6, and 7.
Activity 4.9	Several of these scenarios reflect various of the Keys.				

Cross-Reference Standards for Teacher Competence in Educational Assessment of Students: *Toolkit*

This document cross-references the *Toolkit* to the *Standards for Teacher Competence in Educational Assessment of Students*, which can be used as an outline for teacher education courses.

Text	Standard 1—Skill in choosing assessments	Standard 2—Skill in developing assessments	Standard 3—Skill in administration, scoring, and interpreting results	Standard 4—Skill in using results to make decisions	Standard 5—Skill in grading	Standard 6—Skill in communicating assessment results	Standard 5—Skill in recognizing inappropriate uses of assessment
Chapter 1 Text	Discusses nature of quality.						
Act. 1.1, Handout 1.1	Page 6, lines 1, 3, 4, and 7. Emphasize purposes for assessing students and implications for assessment design.					Page 6, lines 5, 6, and 8. Focus on alternatives in reporting.	Page 6, lines 1, 2, 3, 4, 5, 6, 7, and 8. Designing to eliminate negative consequences.
Activity 1.2							
Activity 1.3						Emphasizes vocabulary.	
Activity 1.4	Designing assessments to help students learn, purpose-method match.						
Activity 1.5	Defining quality assessment, matching assessment purposes to assessment design.		Quality rubrics and scoring performance assessments.	Integrating assessment and instruction.		Illustrates quality communication.	Illustrates potential negative consequences.
Activity 1.6	Methods depend on what and why you want to assess.						
Activity 1.7							
Activity 1.8	Whether the preferred assessment method would change depending on purpose, method-target—purpose match, and multiple measures.			Direct practice on this standard.			Emphasized.
Activity 1.9	Emphasizes clear purposes in choosing the rubric type.		Direct practice on this standard.				

Text	Standard 1— Skill in choosing assessments	Standard 2— Skill in developing assessments	Standard 3— Skill in administration, scoring, and interpreting results	Standard 4— Skill in using results to make decisions	Standard 5— Skill in grading	Standard 6— Skill in communicating assessment results	Standard 5— Skill in recognizing inappropriate uses of assessment
Activity 1.10	Self-assess understanding of quality, purposes, use in instruction, and others.					Self assess understanding of quality, purposes, use in instruction, etc.	
Activity 1.11	The assessment standards in this activity deal with quality—the relationship between complex learning targets for students and methods; using assessment to improve student learning; how assessment relates to instruction; etc.						
Activity 1.12, Handout 1	Page 7, A1, A2, A4, A5, A6, A7, A8. A9.					Page 7, A8.	Page 7, A3, A7.
Chapter 2 Text	Using assessment as a tool for learning and monitoring of student progress.					Implications for communication.	
Activity 2.1	Purposes and relationship to performance rubric design.		Direct practice on this skill.	Implications for this skill.		Implications for communication.	
Activity 2.2	See Act. 2.1.					See Act. 2.1	
Activity 2.3	Illustrates integrating assessment and instruction.					Illustrates communication.	
Activity 2.4		Developing assessments and rubrics.					
Activity 2.5	Performance criteria as a tool for communication to help students learn.						
Chapter 3 Text	Nature of quality in performance assessment: design and purposes.						
Activity 3.1	Nature of quality performance tasks.						
Activity 3.2		Developing assessments.					
Activity 3.3	Nature of quality performance criteria.						
Activity 3.4		Developing assessments.					
Activity 3.5							
Activity 3.6	Consider all areas in order to critique an assessment.						
Chapter 4 Text					Grading and reporting.		
Activity 4.1					Designing grading tools.		

Activity 4.2					Issues in grading and reporting.	
Activity 4.3					Issues in grading and reporting.	
Activity 4.4						
Activity 4.5					Reporting formats.	
Activity 4.6					Converting rubric scores to grades.	
Activity 4.7					Case study in grading and reporting.	
Activity 4.8					Guidelines for grading.	
Activity 4.9					Grading case studies.	

III. Tools to Help You Plan Your *Toolkit* Training

Guide for Planning Your *Toolkit* Assessment Training

This section will guide you through considerations for planning a *Toolkit* training. Whether you are contemplating a trainer-of-trainer session, a long-term districtwide initiative, or a simple 2-hour workshop, give careful considerations to the components listed below and any questions that arise from the them.

Components to Consider	Notes pertinent to planning . . .
1. What is the purpose for this assessment training? The training will be a: ❑ Training of trainers ❑ District initiative ❑ Conference presentation ❑ Teacher training ❑ Community session ❑ Requested session around a specific assessment topic ❑ Other This session level will be: ❑ Awareness session ❑ Follow-up session ❑ Part of a long-term training sequence ❑ One-shot session	
2. What are the goals or major objectives for this training? What is the reason the training is being conducted? ❑ See example sheets in Section 4 for training goals/objectives	
3. Who is your targeted audience? ❑ All educators ❑ Administrators ❑ Mixed teacher group (K-12, all content, etc.) ❑ Grade-level groups of teachers ❑ Content-specific groups ❑ Administrative staff ❑ Other	
4. How much time will be allotted for this training? Hours _____ Day/s _____Multiple dates _____	List dates:
5. Is the training supported by key staff at the school or district level (principal or other key staff)?	
6. Where will the training take place? ❑ If the training takes place at the school where teachers work–how will you assure an uninterrupted session? ❑ Does this site provide a quality learning environment? Is it convenient (climate-controlled, easily located, comfortable, safe, and other)?	Place:

Components to Consider	Notes pertinent to planning...
7. Who is the contact person/s for your training? Name: Address: Phone Number: E-mail:	
8. Is a needs assessment of the targeted audience warranted for the training or for specific assessment professional development activities to be planned? ❑ Yes ❑ No If yes, what type of assessment might be conducted, how will you collect the data, what data do you need, when would it be collected, who would collect it, who would analyze it and report the findings?	
9. What will the agenda include? ❑ Create an agenda (with notes, etc.) to use for you, the professional developer as your guide (see examples in Section 4). ❑ Create an agenda for participants that gives the time, date, activity, or other items (see example sheet).	
10. What are the objectives for participants (expectations, outcomes)? See examples in Section 4 on possible participants' expectations. ❑ If a continuous session is part of this training, what might be the assignment participants will have to do over time? ❑ What products will be produced?	
11. What do you need to consider for organizing the training? ❑ Materials ❑ Handouts ❑ Evaluation forms (see examples in Section 4) ❑ Set up (grouping of participants) ❑ Participants list (to complete and to share with others)	
12. What do you need to review prior to training? ❑ Reading chapters from the *Toolkit* ❑ Reviewing instructions for all activity prior to delivery ❑ Web sites, etc. ❑ Reading in other texts	
13. What type of supplements, handouts, or materials do you plan to provide participants? How much is enough?	

Agenda–4-Day Intensive Training

Day One	
8:00 to-9:00 a.m.	Registration, Coffee, and Conversation Welcome, Introductions: What Tool Are You? Training Purposes IRA and the CBAM *Toolkit* Orientation Activity 1.3 *Post-it Notes–Understanding the Language of Assessment*
11:45a.m. to 12:30 p.m.	Lunch
12:30 to 3:30 p.m.	Five Keys to Quality Assessment Activity 1.5 *Clapping Hands* or Activity 1.7 *Target-Method-Match* Chapter 1 Activities Jigsaw Reflections on the Day, QUICKIES
3:30 p.m.	Evaluation and Closure
Day Two	
8:00 to 8:30 a.m.	Coffee and Conversation QUICKIES, Housekeeping, Agenda Review, and others. Chapter 2 Overview
9:45 to 10:00 a.m.	Break
10:00 to11:50 a.m.	Activity 2.3 *Ms. Toliver's Mathematics Class* Activity 2.5 *How Knowledge of Performance Criteria Affects Performance* Two Hats Reflection
11:50 a.m. to 1:00 p.m.	Lunch
1:00 to 2:00 p.m.	Activity 2.1 *Sorting Student Work*
2:00 to 2:15 p.m.	Break
2:15 to 3:30 p.m.	Next Steps and Planning Time
3:30 p.m.	Reflections and Evaluation
Day Three	
8:00 to 8:30 a.m.	Registration, Coffee, and Conversation
8:30 to 8:40 a.m.	Overview of training purposes and agenda *Post-A-Note* Reflection on assessment activity progress
8:40 to 9:00 a.m.	Chapter 3 Overview
9:00 to 10:15 a.m.	Activity 3.2 *Spectrum of Assessment* Activity, Part B (using PA standards)
10:15 to 10:30 a.m.	Break
10:30 to 11:00 a.m.	Activity 3.4 *The Student's Toolbox*
11:00 a.m. to 12:00 p.m.	Activity 3.6 *How to Critique an Assessment* (Appendix A math and social studies samples)

12:00 to 1:00 p.m.	Lunch
1:00 to 2:00 p.m.	Sharing Progress and Products, in Trios, Part I
2:00 to 2:15 p.m.	Break
2:15 to 3:15 p.m.	Sharing Progress and Products, in Trios, Part II
3:15 to 3:30 p.m.	Interim evaluation, QUICKIES, Closure

Day *Four*

8:00 to 8:30 a.m.	Coffee and Conversation QUICKIES, Housekeeping, Agenda Review, etc.
8:30 to 8:45 a.m.	Chapter 4 Overview
8:45 to 9:45 a.m.	Sharing Progress and Products, Trios, Part III
9:45 to10:00 a.m.	Break
10:00 to10:45 a.m.	Activity 4.2 *Putting Grading and Reporting Questions In Perspective*
10:45 a.m. to 12:00 p.m.	Activity 4.6 *How to Convert Rubric Scores to Grades*
12:00 to 1:00 p.m.	Lunch
1:00 to 2:30 p.m.	Activity 4.8 *Guidelines for Grading*
2:30 to 3:00 p.m.	Table Top Discussion: Changing Assessment Policies and Practices, Barriers and Supports
3:00 to 3:15 p.m.	Next Steps Testimony
3:15 to 3:30 p.m.	QUICKIES, Evaluation

Agenda–1-Day Overview of *Toolkit*

8:00 to 8:30 a.m. **COFFEE AND CONVERSATION**

I. *Toolkit* 98 Introduction, purposes, organization, objectives, participant expectations

II. Overview of Chapter 1, *Standards-Based Assessment, Five Keys to Quality Student Assessment*

Clapping Hands (Activity 1.5)—You'll be the assessors and assessees in this awareness activity about the meaning of quality in assessments and the importance of knowledge of standards and results.

Jigsaw Discussion of Chapter 1 Activities

III. Overview of Chapter 2, *Integrating Assessment with Instruction*

Sorting Student Work (Activity 2.1)—You'll develop a rubric from "back to front" as we sort student work samples on a performance task, experiencing how to develop performance criteria and how to help teachers and students do the same.

Jigsaw Discussion of Chapter 2 Activities

12:30 p.m. **LUNCH (Delivered)**

IV. Overview of Chapter 3, *Designing High-Quality Assessments*

Assessing Learning: The Student's Toolbox (Activity 3.4)—You'll explore different assessment designs for the same learning target and discuss how to "open up" traditional classroom assessments.

Jigsaw Discussion of Chapter 3 Activities

V. Overview of Chapter 4, *Grading and Reporting—A Closer Look*

Grading Scenarios (Activity 4.9)—In pairs, you'll consider some real-life grading dilemmas and suggest solutions.

VI. Using *Toolkit* Resources to Improve Math and Science Instruction and Assessment—Two Hats Reflection Activity

Questions and Answers, Evaluation

4:30 p.m. **CONCLUSION**

IV. Chapter 1 Replacement Pages

Summary

The rationale for alternative assessment and a description of the way it fits into current efforts to improve student achievement is summarized very nicely by the following quotation:

> *The area of achievement assessment has been undergoing major changes during the past few years. A shift has taken place from what some call a "culture of testing" to a "culture of assessment." A strong emphasis is put on integrating assessment and instruction, on assessing process rather than just products and on evaluating individual progress relative to each student's starting point. The position of the student...has also been changing...to that of an active participant who shares responsibility in the process, practices self-evaluation, reflections, and collaboration and conducts a continuous dialogue with the teachers. The [assessment] task is often interesting, meaningful, authentic and challenging....All these changes are part of a school restructuring process, meant to develop self-motivated and self-regulated learners and intended to make learning a more mindful and meaningful experience which is responsive to individual differences among the learners. This shift reflects an "overall assessment prophecy" which holds that it is no longer possible to consider assessment only as a means of determining which individuals [can adapt] to mainstream educational practice....[Now] rather than requiring individuals to adapt to means of instruction, the desired objective is to adapt the means of instruction to individuals in order to maximize their potential for success....The new assessment alternatives being developed enhance the possibilities for adaptation....* (Birenbaum and Douchy, 1996, p. xiii)

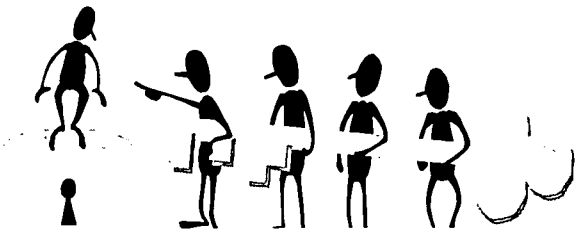
Reference Box:

Menucha Birenbaum and Filip Douchy, 1996. *Alternative in Assessment of Achievements, Learning Processes and Prior Knowledge*. Kluwer Academic Publishers, (781) 871-6600.

The Need for Quality

Given the important role assessment plays in education (and educational reform efforts), it behooves everyone to make sure that assessments are of high quality. It's perfectly possible to design poor quality alternative assessments and use the results for the wrong purposes. Therefore, we're going to back up a little and talk about what good assessment looks like in general. We'll consider all forms of assessment, not just alternative forms of assessment.

Keys to Quality Student Assessment



Sound assessments at any level from the classroom to the boardroom[1]:

- **Arise from clear and appropriate student learning targets.** *What* are we, as educators, trying to assess? We all must clearly and completely define achievement expectations, and these must be couched in the best current understanding of the discipline.

- **Serve a focused purpose.** *Why* are we assessing these targets? Who will use the results and what will they be used for?

- **Rely on a proper method.** *How* will we assess the achievement targets? Will these methods accurately reflect the achievement targets?

- **Sample student achievement appropriately.** *How much* will we collect? Can we be confident that results really reflect what a student knows and can do?

Reference Box

For further reading on keys to quality assessment at the classroom level, see: Rick Stiggins, 1997, *Student Centered Classroom Assessment,* pp. 14-17. Prentice-Hall. Phone: (201)236-7000

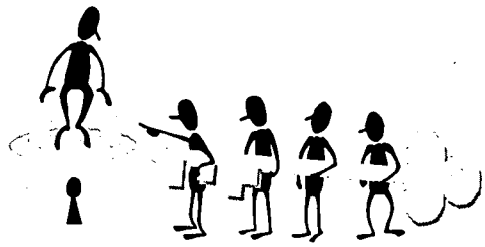
For more information on keys to quality assessment at the large-scale level see: Joan Herman, 1996, "Technical Quality Matters," in Robert Blum and Judy Arter (Eds.), *Student Performance Assessment in an Era of Restructuring*, Section I, Article 7. Association for Supervision and Curriculum Development (ASCD), ISBN 0-87120-267-0. Phone: (800)933-2723

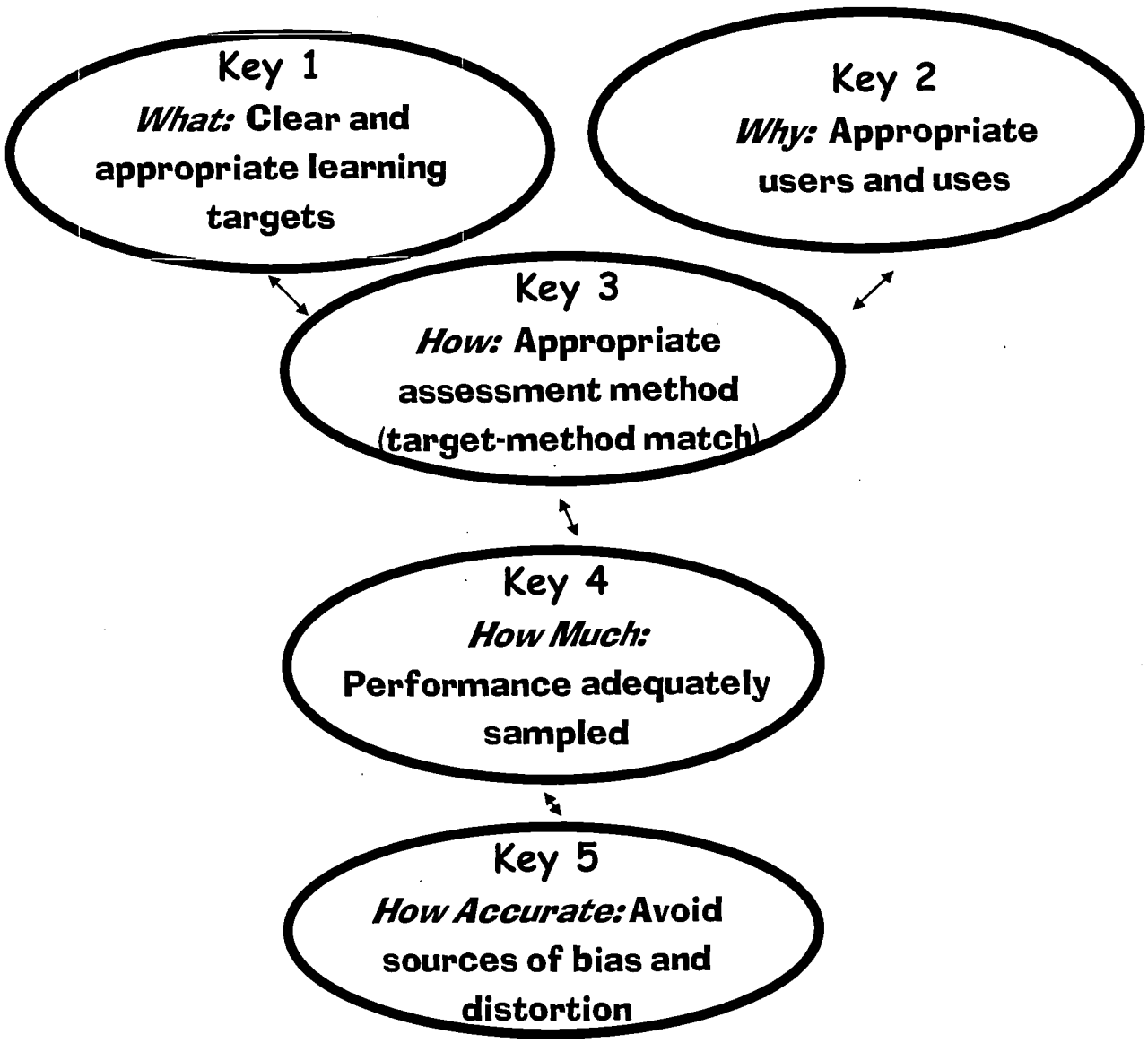
- **Eliminate bias and distortion.** *How accurate?* Did we really assess what we thought we were assessing? Is there anything in the way an achievement target is assessed that masks the true ability of a student or group of students? (For example, too much reading on a math test.)

[1] Based on materials developed by Judy Arter for the Assessment Training Institute, Portland, Oregon. She drew her conceptualization from the two authors listed in the reference box on the right. There is also a cross-reference to the five keys in the *Toolkit* Supplement.

The five keys to quality student assessment are shown in **Figure 1.1** below. The rest of this section will add a little more detail to each key. Chapter 3 will expand on these ideas with respect to alternative assessment.

Figure 1.1 Five Keys to Quality Student Assessment

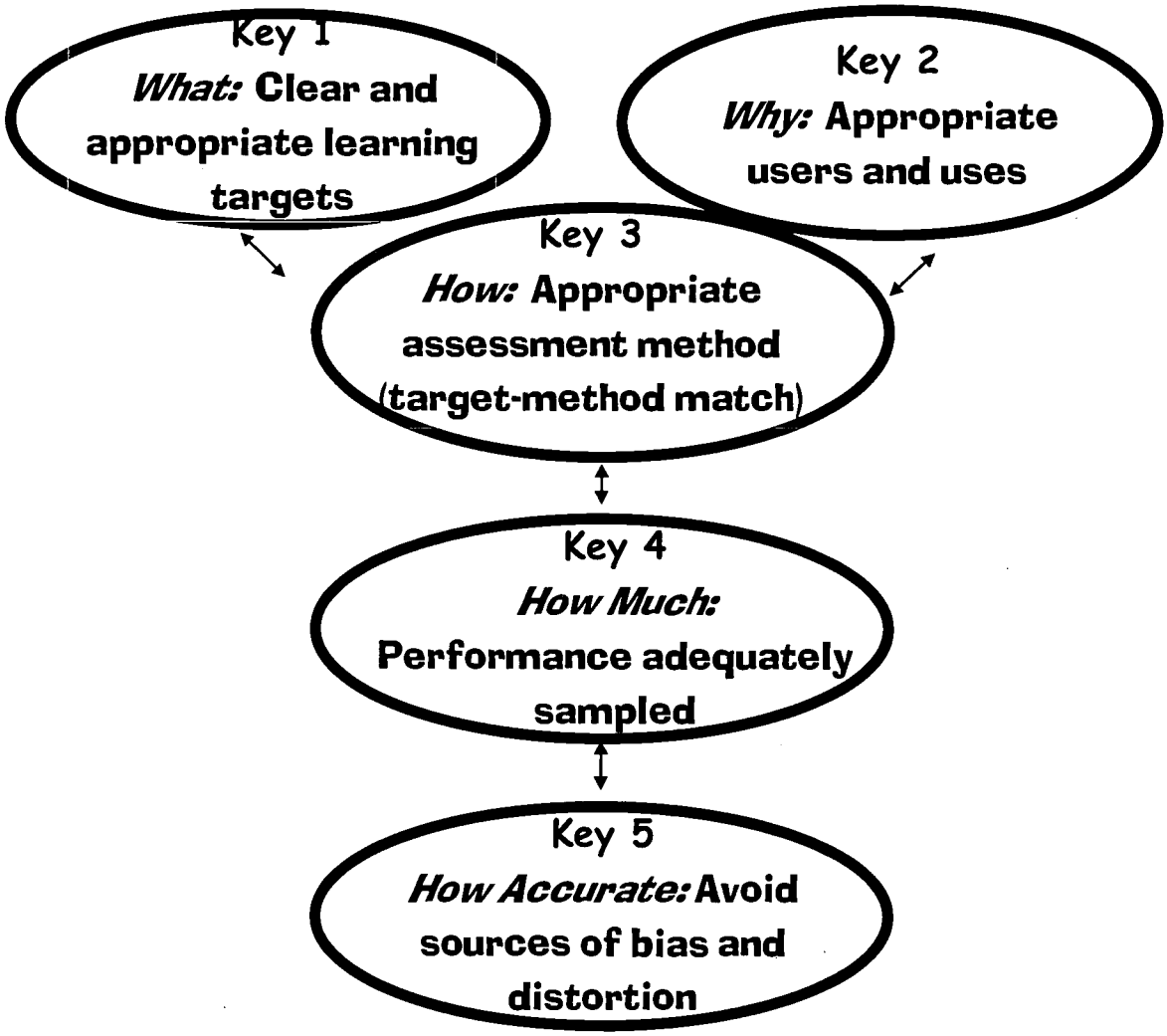


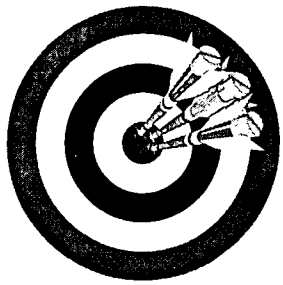


Educators are proficient in assessment to the extent that they can distinguish sound from unsound assessment practices and develop their own assessments using these five standards of assessment quality. These, then, also define what educators need to know and be able to do with respect to assessment—they need to have crystal clear and appropriate learning targets for students, have clear and appropriate purposes for assessing students, use the best assessment method given the purpose and targets, sample student performance appropriately, and, as much as possible, eliminate sources of bias and distortion.

Just as merely listing desired learning outcomes for students is not enough to ensure that everyone understands what they mean, listing the above assessment competencies for educators does not automatically imply how to build competence in these skills. Therefore, the following pages are devoted to describing these skills in more detail—we're trying to practice what we preach about having clear and appropriate learning targets.

Key 1: Clear and Appropriate Learning Targets





The first key to quality with respect to assessment is to have clear and appropriate learning targets. One can't assess something if one doesn't know what it is he or she is trying to assess.

For Key 1, we'll tackle three related topics. The first is "content standards." Standards-setting activities have blitzed the country over the past eight years and are intended to define the "appropriate" part of the "clear and appropriate student learning targets" equation. So, educators need to know about them. Secondly, we'll address the "clear" part of the equation. Then, finally, we'll present a couple of ways to categorize learning targets, and discuss why we might want to classify them.

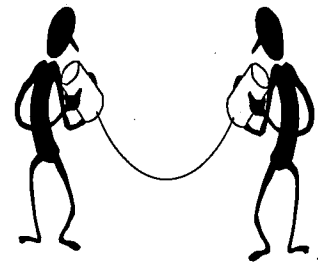
Appropriate Targets/Content Standards. As we've mentioned previously in this chapter, content standards are statements of what should be taught; they specify the "what" of what students should know and be able to do. Content standards come by many names—benchmarks, outcomes, essential academic learning requirements, skills standards, competencies, common curriculum goals, and academic student expectations. Here are some examples:

- Oregon, Grades 6-8 Reading Standard: *Demonstrates inferential comprehension of a variety of printed materials.* Related Grade 8 Benchmark: *Identify relationships, images, patterns or symbols and draw conclusions about their meaning.* ("By Grade Level Common Curriculum Goals, Grades 6-8 Content and Performance Standards," Oregon Department of Education, August, 1996)

- Washington, Writing Standard, Grades 4-10: *The student writes clearly and effectively.* This standard includes the following "components": *develop concept and design, use style appropriate to the audience and purpose (voice, word choice, and sentence fluency),* and *apply writing conventions.* (Essential Academic Learning Requirements, Washington State Commission on Student Learning, January 1997)

- National Standards for Business Education, Secondary: *Demonstrate interpersonal, teamwork, and leadership skills necessary to function in multicultural business settings.* (National Standards for Business Education, National Business Education Association, 1995)

"Nothing new," we hear you thinking, and in a very real sense you're right. The idea of focusing instruction and assessment on that which is most important and enduring is not new. The hard part of it is coming to agreement on what is important and enduring. Have you ever had disagreements with your colleagues on the most important things to emphasize in instruction? Expand this a thousand-fold as a nationwide effort, and you get the picture. Some of the efforts to set standards read like soap operas. But this only underscores the necessity to do it. How can we hold students responsible for outcomes for which we disagree on the meaning? We all owe it to our students and ourselves to be crystal clear on our goals and expectations—no surprises and no excuses.



Clear Targets. Learning targets for students not only need to be appropriate, they need to be clear. It's easy to agree on a target like "communicates well." But, what does this mean? What type of communication, in what contexts, for which purposes? The key to effective student learning targets, be they at the national, state, district, or classroom level, is that they are specific enough to enable everyone to share the same understanding of what students need to know and be able to do. When targets are ambiguous, instruction can take students to vastly different places and assessments can be vastly different. The goal here is *not* to standardize instruction; rather, the goal is to aim at the same learning target even if teachers have different instructional designs.

Learning targets also need to be clear enough so that the persons who find or write assessment items and tasks have the same interpretation of what should be covered as the persons who wrote the target statements. Will future teachers interpret them the same? Are the interpretations clear enough so that, when the assessment results are used to profile achievement strengths and weaknesses, users will know what to do about it? For example, how would one design an assessment for "communicates well"? The assessment could be anything from writing an essay to observing students on the playground as they informally communicate with their peers.

Targets can be unclear for lots of different reasons. In our experience when trying to design assessments to match ambiguous content standards, we have noted the following sources of confusion:

1. Is it clear what cognitive level is being assessed? Is it recall of facts? Higher-order thinking? For example, the word "know" in targets can be interpreted in many ways. How will students act when they "know"? Rote recall of facts, like restating definitions? Or is "knowing" demonstrated through picking out new examples (or counter examples) of the principle, restating the principle in one's own words, or independently using the principle when engaged in real-life tasks?

> **Caution** . . . striking a balance between detail and over-restrictiveness can be tricky. We want targets clear enough that we can all agree on what student success looks like, but not so detailed that "these 10 things" are all we mean. Or, even worse, that "these 300 things" are exactly what we mean.

2. How will assessment of targets repeated across subject areas be handled? Common repeated targets are group skills, oral presentations, critical thinking, study skills, interpreting and using graphs, and so forth. Is there a fundamental difference in how these repeated targets are expressed in different subjects? If so, how will the assessments be different? If not, how will the assessments be the same? How will instruction and assessment complement each other to give a full picture of all aspects of repeated targets? Should repeated targets be assessed as part of subject area assessments, or separately?

3. Are there any holes in the statement of essential learning targets for students? Are there important targets not covered anywhere? Do targets reflect current thinking about what expertise in a content area looks like? A standards-based assessment can be only as good as the target statements on which it is based.

Rule of thumb: *Is the target clear enough that a group of teachers would agree on the range of knowledge, skills, and performance implied by the target? Would they agree on what to teach and what to assess?*

Tricky? You bet. In fact, Joan Herman and her colleagues at UCLA state that available evidence suggests that many states' standards currently are not strong enough to support rigorous assessment development. *But, we believe that it's the attempt to clarify targets, as much as having final clear targets in place, that makes a difference.* In groups we've worked with, the general consensus is that everyone who makes the effort ends up with a much more indepth understanding of what they are trying to accomplish with students.

Types of Learning Targets. There are a million (well, actually maybe a hundred) different ways to categorize the types of learning targets we've seen for students. But, hold on, you're saying, why would one even want to "categorize" them? Well, this isn't just an outgrowth of compulsiveness on the part of number-crunchers. The process of categorizing helps to do three things. First, it helps folks to thoroughly think through what they want students to know and be able to do (in other words, clarify targets). Second, it helps folks determine if they have a good mix of learning targets. And, finally, it will help folks, later, to choose the appropriate assessment method.

Reference Box

For further reading on types of targets à la Rick Stiggins, see: Rick Stiggins, 1997, *Student Centered Classroom Assessment,* Chapter 3. Prentice-Hall.
Phone: (201) 236-7000.

Types of Learning Targets. There are a million (well, actually maybe a hundred) different ways to categorize the types of learning targets (achievement goals, outcomes) we've seen for students. But, hold on, you're saying, why would one even want to "categorize" them? Well, this isn't just an outgrowth of compulsiveness on the part of number-crunchers. The process of categorizing helps to do three things. First, it helps folks to thoroughly think through what they want students to know and be able to do (in

other words, clarify targets). Second, it helps folks determine if they have a good mix of learning targets. And, finally, it will help folks, later, to choose the appropriate assessment method.

Here is one "take" on how to categorize learning targets for students. The first thing to remember about this (or any other) categorization scheme is that it is a convenience made up by someone in order to help people discuss things that are complex. There is no "truth" out there in the universe that "there are five kinds of student learning targets" and that every student learning target fits neatly into one of these five types.

Rick Stiggins finds that classifying student learning targets into five categories helps teachers find a good mix in instruction and assessment. His five types of student learning targets are:

- **Knowledge Mastery**—Knowing and understanding substantive subject matter content, including facts (e.g., "John Kennedy was assassinated on November 22, 1963"), generalizations (e.g., "People holding high political office put their lives in jeopardy"), and concepts (e.g., "political assassinations").

- **Reasoning Proficiency**—The ability to use content understanding to reason and solve problems. Reasoning includes things such as analyzing, comparing, thinking critically, and decisionmaking.

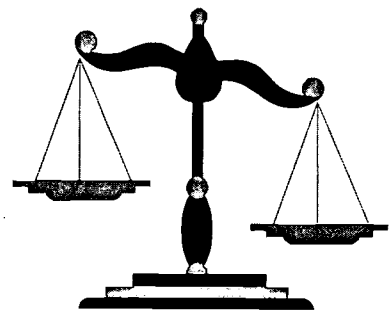
- **Skills**—Doing things such as reading fluently, working productively in a group, making an oral presentation, speaking a foreign language, or designing an experiment.

- **Ability to Create Products**—Creating tangible products such as an essay, a research report, visual art, or a wood table.

- **Dispositions**—Student attitudes, including: attitude toward school, civic responsibility, self-confidence, desire to learn, flexibility, and willingness to cooperate.

Caution . . .

Affective targets can be a red flag in some communities. If so, the user can delete any references to the affective domain in this material and stick just to the cognitive domain.



Reflection questions on **Key 1,** clear and appropriate student learning targets:

1. What are the five most important things you'd like your students to be able to do as the result of your time with them? Do students know this? Is it clear to you and your students how you will know when they achieve these targets?

2. Look at local (state, district, or classroom) content standards (learning targets). Are there any that might need further definition to make them clear? Would all teachers agree on what they mean?

3. Can you identify any learning targets currently assessed that might not be worth the time devoted to them?

Related *Toolkit* Chapters and Activities

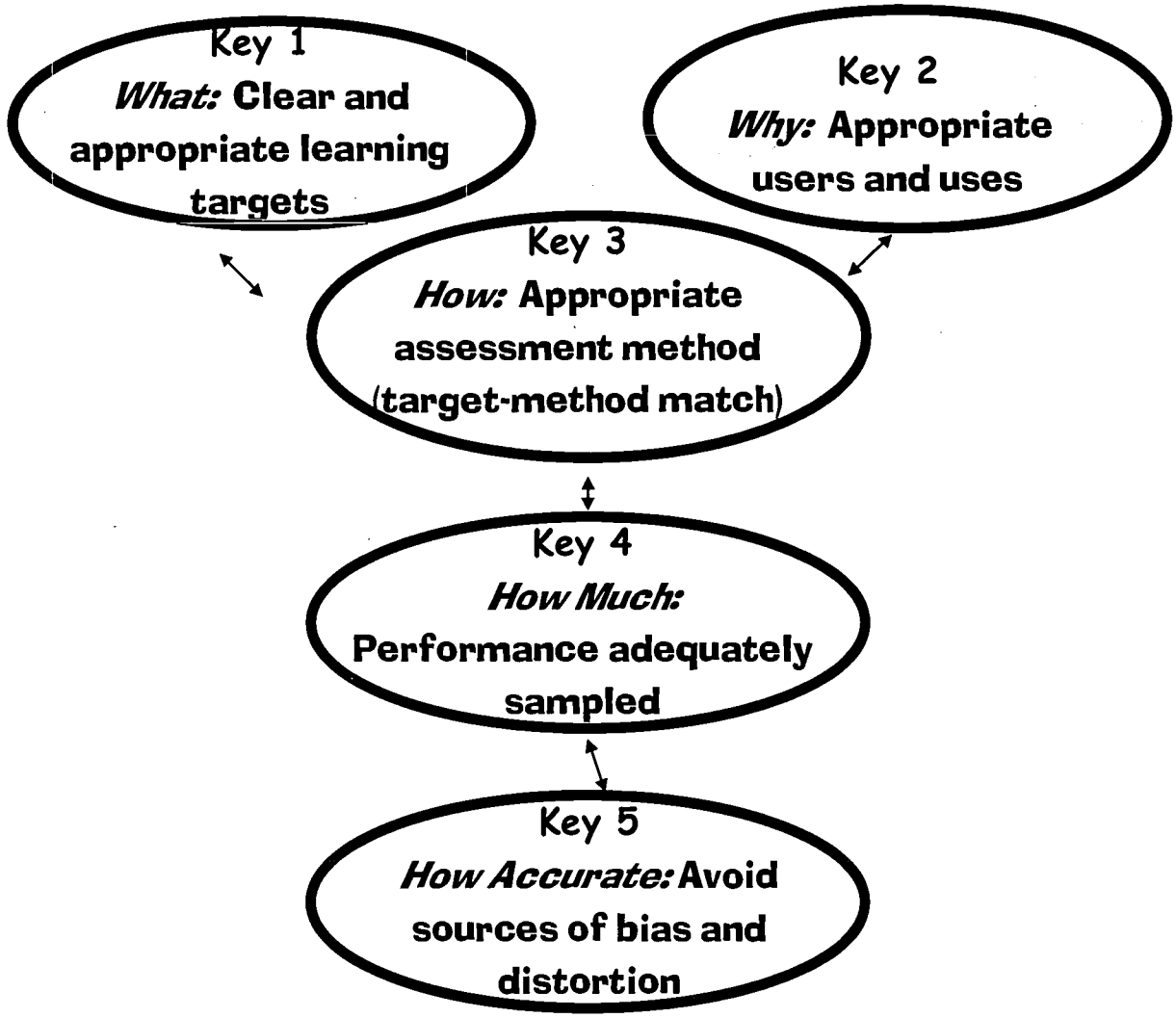
Activity 1.2—*Clear Targets—What Types Are These?* Asking people to describe what kind of target they are looking at (knowledge, reasoning, etc.) is an excellent way to begin to tease out differences in what is meant by target statements.

Activity 1.6—*A Comparison of Multiple-Choice and Alternative Assessment.* Comparing different ways to assess a content area provides a means of exploring what it means to know and understand a content area.

Activity 1.10—*Clear Targets and Appropriate Method—The View From the Classroom.* In this activity, participants are asked to self-evaluate the clarity of their learning targets using a rating form.

Activity 2.1—*Sorting Student Work.* Analyzing what makes student work effective is an excellent way of opening up the discussion of what it means, for example, for a student to write well.

Key 2: Focused and Appropriate Purpose



We'll talk about two things in this section. First, we'll address *why* educators assess—purposes. In other words, who are the users and uses of assessment results? This will underscore the importance of doing a good job of assessing students. Then, we'll look at "the rest of the story"—how purpose actually affects the way an assessment is designed.

So, why do folks assess student achievement? Who uses the results and what do they use the results for? Well, just about everybody for just about everything, and assessment activity seems to get more intense every day. For example:

- **Teachers** assess students day-to-day in the classroom for such purposes as planning instruction, evaluating what worked and what didn't, grading, promoting student self-control of progress, and communicating with parents.

- **Students** use the results of assessments to decide what they'll study, how much they'll study, whether it's even worth studying, what they're good at (or not), their self-worth,

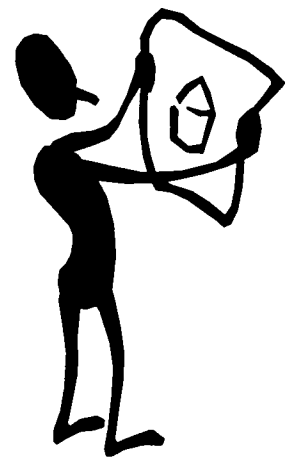
who they might associate with, and how they might earn their living as an adult. Assessment information can also give feedback to students so they can reflect on their performance and make changes. Ultimately, we want students to understand how they learn, evaluate their own performance, and undertake the learning necessary to improve their performance.

- **Parents** use the results of assessments to determine rewards and punishments for their children, apportion family resources (who will go to college?), make judgments of family self-worth, decide whether to vote for a levy, and pick a place to live.

- **Principals and other administrators** use assessment results to promote and graduate students, allocate resources, plan professional development, and report to the public how well the school or district is doing.

- **State departments of education** use assessment results to report to the public, levy rewards and sanctions for districts, and distribute resources.

Looking at this list reminds one that these are pretty important uses for assessment results and that we all had better be darned sure that our assessments are of good quality. What would happen, for example, if an assessment gave an inaccurate picture of student achievement? What would happen if what was *actually* assessed was not really what was *thought* to be assessed? Or worse yet, what if users were unsure as to what they were assessing, so they didn't know what the results really meant? Is everyone *positive* that they are accurately assessing the most enduring outcomes for students so that the decisions made can really serve to guide learning?



Looking at the list of users and uses also reminds one of the crucial importance of not only large-scale assessments—those that occur in roughly the same way at roughly the same time across classrooms—but also *classroom* assessments. After all, which assessments—day-to-day classroom assessments or once-a-year large-scale assessment—most affect the kinds of decisions made by teachers, parents, and students? (We would choose classroom assessment; we hope readers did, too.) What happens if classroom assessments are not well thought out and executed?

The point here is not to suggest that teachers should be blamed for lapses in their knowledge about classroom assessment. After all, most teachers never had the opportunity to learn about assessment because most states don't even require an assessment class for certification. And, even in places where an assessment course is required, there is an evolving understanding of what teachers really need to know and be able to do to be good classroom assessors.

The point *is* that student assessment is of crucial importance. That's why there is activity on all fronts to improve it, from clearly defining valued student learning targets at the state level and rethinking how to best assess them in large-scale assessment, to changing coursework for preservice teachers, to assisting teachers to align important learning targets to instruction and assessment in the classroom. *One of the efforts to assist teachers to fine-tune classroom assessment practice is this Toolkit!*

Now for the second part of this section—**How does purpose affect how educators assess?** It's probably obvious that an assessment good for one purpose—for example, providing detailed diagnosis of a student's strengths and weaknesses—is not necessarily best for other purposes—for example, determining the strengths and weaknesses of the school's overall curriculum or whether most students attained the school's grade-level goals for student performance. Thus, a first major decision to make is deciding one's purpose for assessment.

Just consider the differences in uses of information from large-scale and classroom assessments. In general, large-scale purposes require more rigorous evidence of technical quality than do classroom assessments, primarily because important decisions are likely to be based on them and because it is usually only a single testing episode. In contrast, for classroom purposes, a teacher has lots of formal and informal evidence upon which to base decisions, and so the results of any single, faulty assessment are not likely to be given undue weight.

As other examples of how purpose can affect assessment design, consider these:

1. A single multiple-choice or short answer multiplication test may be perfectly acceptable to determine whether or not third-graders have learned their multiplication facts, but would not be appropriate for making a decision about the overall quality of the third-grade mathematics program.

2. A short answer or multiple-choice assessment designed to measure student knowledge of specific scientific principles might be useful for partially determining a student's grade. Inferring that this assessment sufficiently measures the student's ability to perform scientific tasks that call for an understanding of these principles would require the assessors to observe the student applying that scientific knowledge in a laboratory setting.

3. Some assessments are used mostly to gather information about students in order to make decisions about them—for example, grading or certifying competence. Other assessments are designed more to involve students in their own assessment and thus serve an instructional function. This distinction between assessment to "monitor" and assessment to "teach" has implications for assessment design. For example, rubrics used by teachers to monitor student performance might not need to be as detailed as rubrics used by students to learn and practice the features of writing that make it work.

Reflection questions on **Key 2,** clear and appropriate users and uses:

1. What other examples can you think of where the purpose of the assessment affects how the assessment occurs? Selection for special programs? Performance assessment in the classroom versus in a large-scale context?

2. Why do *you* assess students? Is this the same or different from the purposes of your colleagues? Is there life (assessment) beyond grading?

3. What are the occasions when you engage in assessment activities solely to build student self-assessment skills? How do these occasions differ, if at all, from those in which you're assessing students to find out what they know and can do? What student skills are built through involving students in their own assessment?

4. When have you altered how you presented student achievement results because of the audience? How did you ensure that different audiences could understand and use the results?

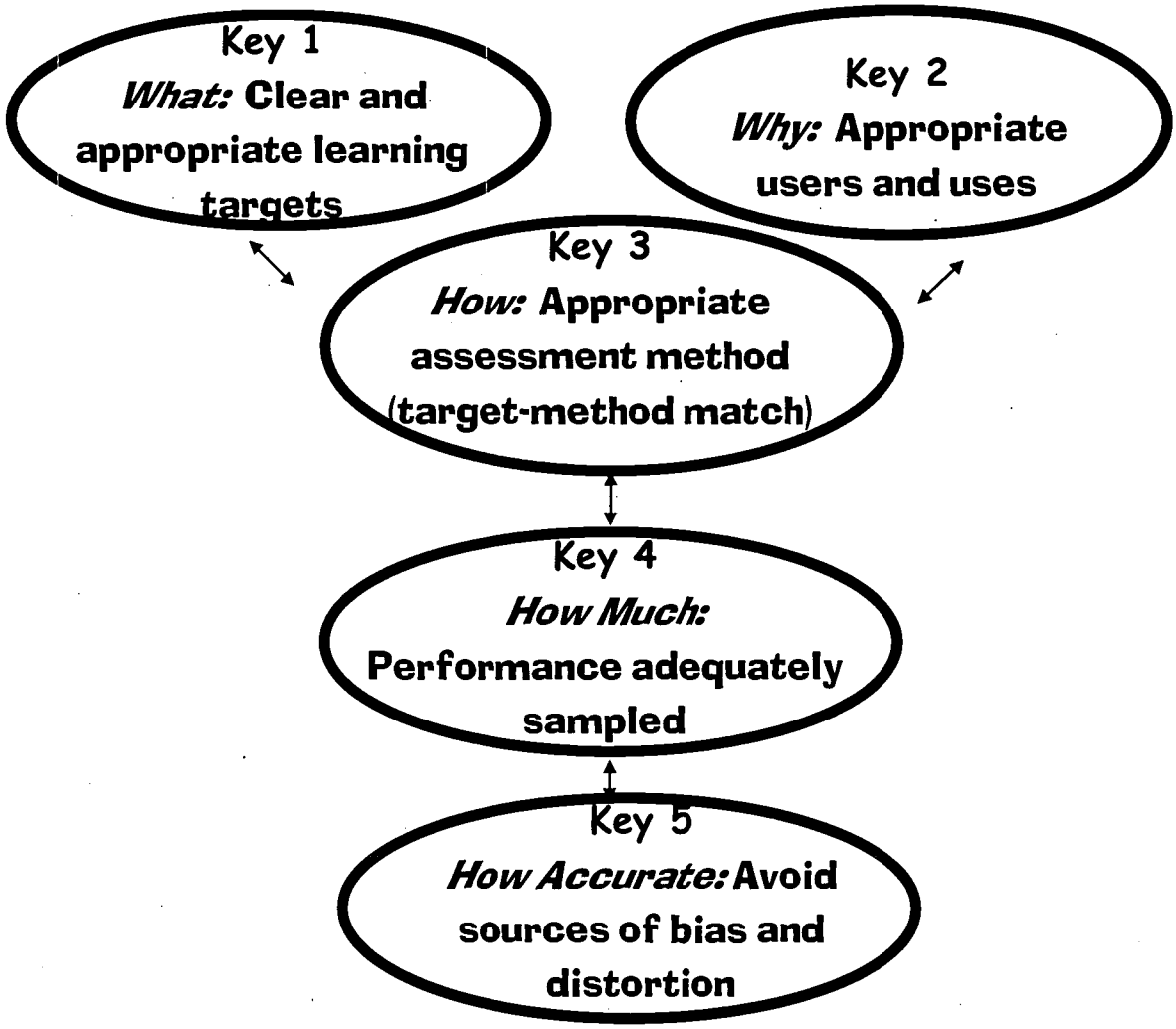
Related *Toolkit* Chapters and Activities

Activity 1.8—*Sam's Story* asks participants to judge the assessments they'd trust to give good information for a particular purpose—determining proficiency in math for instructional planning. At the end, it asks whether other purposes (for example, whether a student is working up to potential) might require different assessments.

Activity 1.9—*Going to School* asks participants to think about different designs for performance criteria and which might be most useful for different purposes.

Chapter 2 addresses various purposes for classroom assessment and posits design implications.

Key 3: Appropriate Methods (Target-Method Match)



The third key to quality assessment is to match targets and purposes to methods. We are *not* of the opinion that the only good assessment is a performance assessment. Rather, there are times and places for all different forms of assessment—multiple-choice, matching, true/false, essay, performance assessment, portfolios, and oral communication. Good assessment means having a clear idea of what we want to assess and then picking the best way to assess it.

We'll use Rick Stiggins' "target-method match," ideas. Again, there is no "truth in the universe" that says, for example, all skills targets should be assessed with a performance assessment. Stiggins'scheme is merely one person's attempt to assist us to order our thinking about assessment topics that are complex. You will be able to think of exceptions to Stiggins' This is good! The goal is to have defensible reasons for choosing the methods used.

Stiggins maintains that, although one can assess most types of student learning targets by most methods, there are some more and less efficient ways to do it. For example, if all you want to find out is whether students know their multiplication facts, why design performance assessments? **Figure 1** shows his recommendations for matching targets to methods. X's denote a good match; O's denote a partial match.

Figure 1.3
Aligning Achievement Targets and Assessment Methods

	Selected Response	Essay	Performance Assessment	Personal/Oral Communication
Knowledge Mastery	X	X		O
Reasoning Proficiency	O	X	X	X
Skills			X	X
Products		O	X	X
Dispositions	X	O	O	X

X = good match; O = partial match

Now, the rationale for the X's and O's: It's simply not efficient to use performance assessment or personal oral communication to assess every *knowledge* outcome educators have for students. For example, using performance assessments to see whether students know all their multiplication facts could take years. But we *could* assess instances of ability to multiply in the context of a problem-solving performance task.

While it is possible to assess some kinds of student *reasoning* skills in, say, multiple-choice format, to really see reasoning in action one needs a more complex assessment format. For example, most standardized, norm-referenced tests have questions about such things as fact versus option and "what is most likely to happen next." But these are usually assessed out of context as a discrete skill. One would need a performance assessment to see how students can use all their reasoning skills together to address an issue, or to see if they know when, for example, they need to identify an opinion.

Knowledge *about* what it takes to *perform skillfully* or to *produce a product* can be assessed in multiple-choice format, but to actually *see* if a student can *do* it, one needs a performance assessment. (For example, one can assess student knowledge about how to give a good oral presentation through an essay, but if one wants to see if students can apply this knowledge, one has to have students actually *give* an oral presentation.)

Selected response questionnaires can tap student *dispositions*, but so can open-ended questions (essays) and personal communication with students.

Reflection questions on **Key 3,** matching methods to targets and purposes:

1. Are there any of these assessment methods that you'd like to know more about before you feel confident in choosing the best method for given targets and uses?

2. Do you agree with Stiggins' matching table? Why or why not?

3. Can you think of specific examples that would fit in each cell of Stiggins' table?

4. Where might you consider changing methods in order to better assess a target? Do you use a good mix of methods? Are there any methods that you use very infrequently?

Related *Toolkit* Chapters and Activities

Activity 1.6—*A Comparison of Multiple-Choice and Alternative Assessment* provides participants an opportunity to compare a traditional multiple-choice assessment with a performance assessment and discuss when each should be used.

Activity 1.7—*Target-Method Match* introduces assessment methods and gives participants practice in matching methods to learning targets.

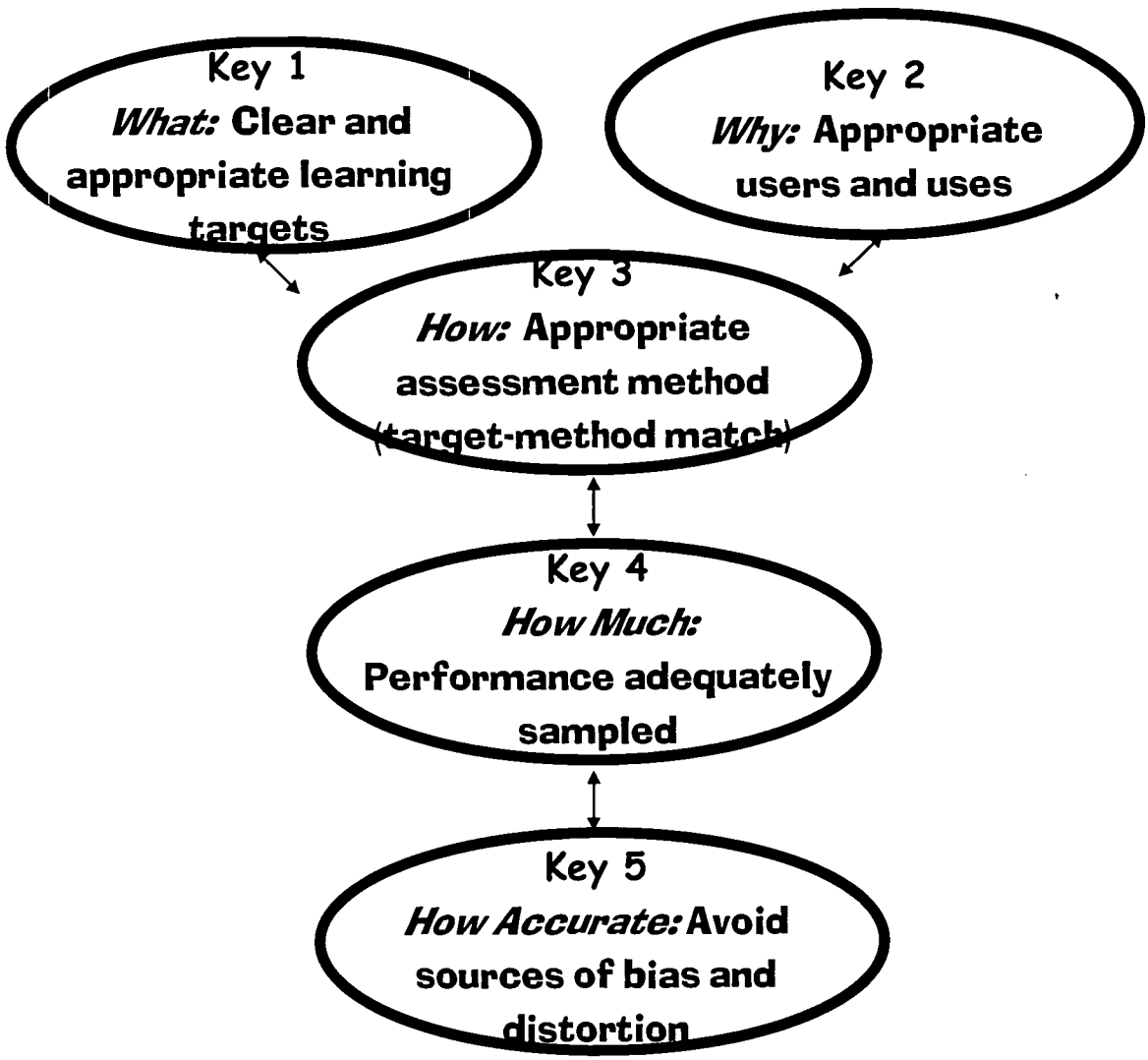
Activity 1.10—*Clear Targets and Appropriate Method—The View From the Classroom* asks participants to self-evaluate the extent to which they successfully match assessment methods to targets.

Chapter 3 covers design options for alternative assessment and includes additional discussions of target-method match.

Activity 3.2—*Spectrum of Assessment Activity* looks at how to "open up" traditional assessment tasks in order to measure more complex outcomes.

Activity 3.4—*Assessing Learning: The Student's Toolbox* demonstrates the relationship between assessment tasks and the student learning that is intended to be assessed.

Key 4: Performance Adquately Sampled



How much information about student achievement is enough? Examples are worth a thousand words in this case. So, here are two examples that illustrate the issues involved in sampling.

Example 1—Certifying Competence in Writing. The state of Confusion is trying to determine individual student writing proficiency so that it can certify student competence. Its procedure is to administer a multiple-choice test that covers spelling, grammar, usage (and all the other stuff that multiple-choice tests cover), and a writing sample that asks students to compare and contrast two pieces of literature. Does this give sufficient evidence of competency to write well?

We, of course, are hoping that you're screaming, "No, no, no!" Confusion's writing assessment doesn't obtain a good, representative sample of various types of writing for different audiences and purposes. Also, the particular topic assigned to the student might not invite each student's best effort. And, some students might just not have been at their best that day for whatever reason. To get a good, stable estimate of ability to write, one would need to gather six or so samples of writing for different audiences and purposes. The state of Confusion might want to try a portfolio system.

This example illustrates the need to adequately sample from the complete range of the skills and knowledge a target implies. (This range of knowledge and skills is called a "domain." So, one hears about the need to "sample the domain well.") It also illustrates the need to collect enough samples so that one obtains a stable estimate of student performance.

Example 2—Impressing the School Board or Seeing How Students are Doing? So, the school board in Elbow Bend District wants to see how well students are learning health. Right now there is no mandated time-per-day requirement for teaching health, and the school board is considering whether to institute such a policy. To keep teacher anxiety levels down, district staff decide to ask selected teachers in each school to provide samples of student work that illustrate how much students know about health.

The result? They got the best examples from the best students in the classes of the best teachers. With their procedures, they did not answer the question, "What is the *typical* student learning in health?" Rather, their sampling procedure led them to answer the question, "What are the *best* students in the *best* classes learning about health?" It's OK to answer the latter question, it's just not OK to say that one has answered the former question when they collected information in the latter manner.

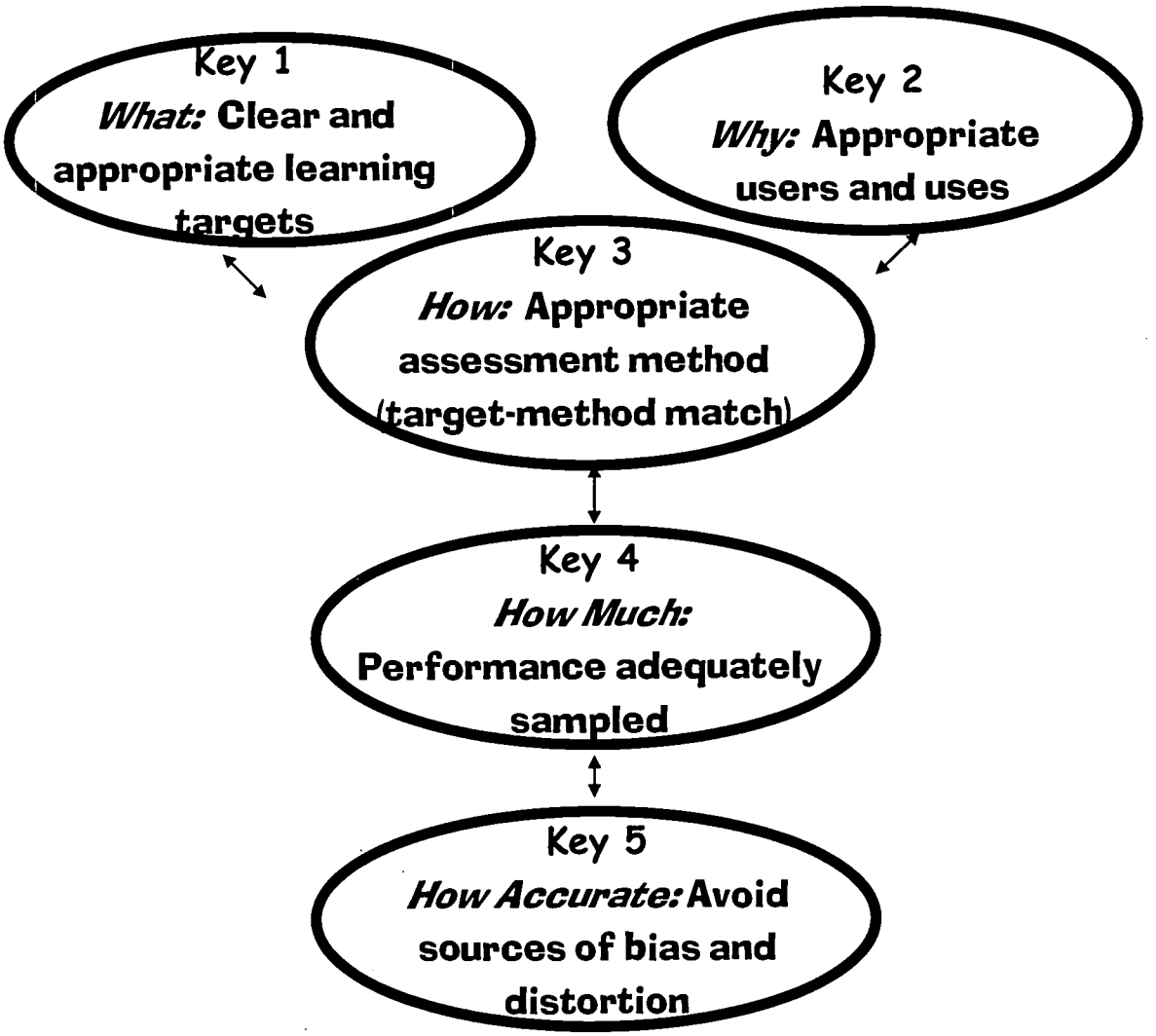
The point of this example is that the sampling procedure has to match the question to be answered. Best work? Typical work? And for whom— *all* fourth-graders as a *group* or *individual students*? Each of these questions implies a different sampling plan.

Reflection questions on **Key 4,** sampling:

1. Can you remember any time when you felt uncomfortable because a decision was made based on too little data?

2. Pick a student learning target, such as "reads with comprehension." Develop a plan for assessment that would provide adequate information upon which to decide when this occurs. What is the breadth and depth of the range of this target (its domain)? What types of reading? What types of responses? How many samples?

Key 5: Eliminating Sources of Bias and Distortion

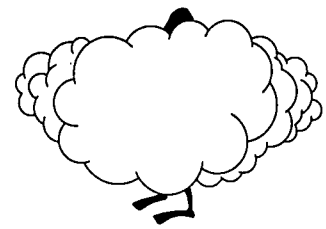


Assume that targets are perfectly clear and appropriate, purposes have universal agreement, the absolutely best way to assess each target has been picked, and you've planned for an adequate sample of performance. Is this all? Unfortunately, no. It's still possible, even easy, to execute the plan poorly.

In short, things can go wrong in assessment. Have you ever tried to engage students in an instructional activity and not gotten at all what you expected? The instructions weren't clear, or there wasn't enough time, or students didn't have all the prerequisite skills, or the activity didn't allow students with different learning styles to do their best.

Well, the same thing can happen in assessment. These "things that go wrong" are called sources of mismeasurement, bias and distortion, or invalidity. The result is that the information from the assessment doesn't mean what we think it means.

What happens if the ability to read the instructions interferes with a student's ability to demonstrate math skills? Or, the necessity to write a response interferes with how well a student can demonstrate ability to set up a scientific experiment? Then these assessments could really measure reading or writing rather than math or science—these are serious potential sources of bias and distortion.



If these assessments then form the basis for a grade or for certifying competence on a state graduation test, the result would be unfortunate. A grade or certification of competence is only as good as the assessment upon which it is based.

This, then, is the fifth key to quality—attending to what might go wrong and fixing it as much as possible. You've seen lists of what can go wrong in assessment. But, just for completeness sake, see **Figure 2.**

Figure 2
Possible Sources of Bias and Distortion in Student Assessments*

Sources due to the assessment itself	• Too much reading or writing on an assessment designed to assess something besides reading or writing • Unnecessarily difficult or unfamiliar vocabulary used in instructions • An assessment method that doesn't allow students with different learning styles to do their best • Unclear instructions • Attempts to make a problem more "real-life" that results in a context more familiar to some groups of students than others • Performance assessments: rater bias; untrained raters; performance criteria that don't cover the most important aspects of performance; performance criteria that are vague; required materials that don't work; tasks that don't really elicit the skill to be assessed • Multiple-choice tests: irrelevant clues to the right answer; more than one right answer; unnecessarily convoluted questions
Sources due to the student	• Student is not rested, is hungry, is sick, or is distracted for some other reason • Student is not used to the format, timing, or other logistics of the testing situation
Sources due to the environment	• Noisy or distracting environment • Assessment administrator that projects a negative attitude toward the assessment • Assessment not given as instructed • Teacher assistance with the assessment
*A sampling	

Reflection questions on **Key 5,** eliminating potential sources of bias and distortion:

1. Did you ever take an assessment you felt didn't really show what you knew or could do? What was it about the assessment that interfered with your ability to shine?

2. Have you ever had a student where you felt that the assessment didn't adequately reflect what he or she knew or could do? What was it about the assessment that interfered with the student's ability to shine?

Related *Toolkit* Chapters and Activities

Eliminating potential sources of bias and distortion is discussed in many ways in many places in *Toolkit.* **Chapter 3** contains an indepth discussion of various alternative assessment designs and the relative merits of each approach. Much of this discussion relates to issues of potential sources of bias and distortion.

Activities that stress the importance of quality include:

Activity 1.5—*Clapping Hands* (potential sources of bias and distortion in performance assessments)

Activity 1.8—*Sam's Story* (the most valid pieces of information for a particular purpose)

Activity 1.11—*Assessment Standards*

Activity 1.12—*Assessment Principles* (equity)

Activity 3.1—*Performance Tasks, Keys to Success* (characteristics of quality tasks)

Activity 3.3—*Performance Criteria, Keys to Success* (characteristics of quality criteria)

Activity 3.6—*How to Critique an Assessment* (practice critiquing on all aspects of quality)

Activity 3.7—*Chickens and Pigs* (equity)

Activity 3.8—*Questions About Culture and Assessment* (equity

Activity 3.9—*Tagalog Math Problem* (equity)

Summary of Steps to Quality

We've had a brief excursion into keys that describe quality assessment in any context, for any purpose, for all types of assessment. The remainder of *Toolkit* focuses on alternative assessment (particularly performance assessment), not because we don't think other forms of assessment are worthwhile, but because that's the area in which most people want assistance right now. So, we'll summarize our look at the Keys in terms of when each occurs in the alternative assessment design process. Many of these steps are covered in more detail throughout *Toolkit*. We've noted where in *Toolkit* each is expanded upon. Some topics are beyond the scope of *Toolkit*. (We have noted those that are not addressed in detail.)

1. Clearly define what it is you want to assess. (This is **Key 1**—*Clear and Appropriate Learning Targets*.) In *Toolkit* we focus primarily on those targets most appropriately assessed by alternative assessment procedures.

2. Clearly state the purpose for the assessment, and don't expect one assessment to work for other purposes for which it was not designed. (This is **Key 2**—*Appropriate Users and Uses*.) In *Toolkit* we focus primarily on classroom purposes for assessment such as monitoring student achievement, planning instruction, and engaging students in self-assessment. These purposes are expanded on in **Chapter 2**—*Integrating Assessment with Instruction*.

3. Match assessment methods to the achievement target(s) and purpose. (This is **Key 3**—*Appropriate Assessment Method—Target-Method Match*.) Developing alternative assessments for classroom assessment purposes is the topic of **Chapter 2**—*Integrating Assessment with Instruction* and **Chapter 3**—*Designing High-Quality Assessments*.

4. Specify illustrative tasks (activities, problems) that would require students to demonstrate the desired skills and accomplishments. Avoid tasks that may only be interesting activities for students, but may not yield evidence of a student's mastery of the desired outcomes. (These are **Key 3**—*Appropriate Assessment Method—Target-Method Match* and **Key 4**—*Performance Adequately Sampled*.)

5. Avoid tasks in which student performance will not come through due to bias. (This is **Key 5**—*Avoid Sources of Bias and Distortion.)* Design options and quality considerations for tasks are in **Chapter 3**—*Designing High-Quality Assessments.*

6. Specify the criteria and standards of judging student performance on the task selected in Step 4. Be as specific as possible, and provide samples of student work that exemplify each of the standards. Develop a reliable rating process that would allow different raters at different times to obtain the same, or nearly the same, result. If used in the classroom by a single teacher, the rating system must allow consistency across students. (These are **Key 3**—*Appropriate Assessment Method—Target-Method Match* and **Key 5**—*Avoiding Sources of Bias and Distortion*.) Criteria and

development of scoring processes are dealt with extensively in **Chapter 2**—*Integrating Assessment with Instruction* and **Chapter 3**—*Designing High-Quality Assessments.*

7. Avoid other pitfalls that can lead to mismeasurement of students. (This is **Key 5**—*Avoid Potential Sources of Bias and Distortion.*) A gentle introduction to technical issues, especially as related to classroom assessment, is found in **Activity 1.5**—*Clapping Hands.* Equity and other quality issues are explored in **Chapter 3**—*Designing High-Quality Assessments.*

8. Collect evidence/data that show that the assessment is reliable (yields consistent results) and valid (yields useful data for the decisions being made). For performance assessments, this might be demonstrated through the level of agreement between scores given by different assessors for the same student work, and evidence that students who perform well on the assessment also perform well on other related items or tasks. This topic is not covered in detail in *Toolkit.*

9. Ensure "consequential validity"—the assessment maximizes positive side effects and minimizes negative ones. For example, the assessment should give teachers and students the right messages about what is important to learn and to teach; it does not restrict the curriculum; it is a useful instructional tool; and the decisions made on the basis of the assessment results are appropriate. Consequential validity and the messages our assessments send teachers, students, and parents are discussed extensively in **Chapter 2**—*Integrating Assessment with Instruction* and **Chapter 3**—*Designing High-Quality Assessments.*

10. Use test results to refine assessment and improve curriculum and instruction; provide feedback to students, parents, and the community. **Chapter 2** and **Chapter 3** further discuss the relationship between assessment and instruction; **Chapter 4**—*Grading and Reporting, A Closer Look* addresses grading and reporting issues.

Chapter Summary and Conclusion

This chapter provides the reader with an overview of the reasons changes are being made in assessment and with a beginning understanding of the issues assessors must grapple with as they use (or develop) assessments to make educational decisions about students and programs. The remaining chapters provide a more indepth discussion of these topics.

Assessment is changing because education is changing. Research shows that what students need to know and to be able to do, and the way in which knowledge and skills need to be taught, are often quite different from conditions in the past. This is not an indictment of schools. Changes in the world are happening at such a rapid pace that few aspects of society have been able to keep up. But keep up we all must, if our students are to be successful in the changing world of tomorrow.

Although we present a strong case for alternative assessment in *Toolkit,* we neither say that all assessments need to be of this type nor reject the use of multiple-choice and other forms of selected-response tests. We do affirm that alternative assessments, when designed well, offer appealing ways to assess complex thinking and problem-solving skills and, because they are grounded in realistic problems, are potentially more motivating and reinforcing for students. However, while alternative assessments may tell us how well and deeply students can apply their knowledge, selected-response (e.g., multiple-choice tests) may be more efficient for determining how well students have acquired the basic facts and concepts. Educators need to choose assessments based on professional information, not on tradition ("that's how I was tested"), personal opinion ("I like them"), or trendiness ("Grant Wiggins says they're great.") A balanced curriculum requires a balanced approach to assessment.

The need for high-quality assessment information to make informed decisions about changes in students and programs will be critical to the success of the educational improvement effort. There is no single correct method for assessing students, but there are ways in which all forms of assessment can be used well to help schools make good decisions and meet students' needs. Keys to quality include: clear and appropriate learning targets; appropriate users and uses; appropriate assessment method (target-method match); performance adequately sampled; and avoiding potential sources of bias and distortion.

TM033483

U.S. DEPARTMENT OF EDUCATION

OFFICE OF EDUCATIONAL RESEARCH AND IMPROVEMENT (OERI)

EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

REPRODUCTION RELEASE (Specific Document)

I. DOCUMENT IDENTIFICATION

Title: ADDENDUM TO IMPROVING CLASSROOM ASSESSMENT: A TOOLKIT FOR PROFESSIONAL DEVELOPERS

Author(s): Regional Educational Laboratories

Corporate Source (if appropriate): Northwest Regional Educational Laboratory

Publication Date: 9/00

II. REPRODUCTION RELEASE

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, Resources in Education (RIE), are usually made available to users in microfiche and paper copy (or microfiche only) and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the options and sign the release below.

CHECK HERE ➧ [XX] Microfiche (4" x 6" film) and paper copy (8½" x 11") reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Northwest Regional Educational Laboratory

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

OR [ ] Microfiche (4" x 6" film) reproduction only

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

_____

_____

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed in both microfiche and paper copy.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction of microfiche by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

SIGN HERE ➧

Signature: Jerry D. Kirkpatrick Printed Name: Jerry D. Kirkpatrick

Organization: Director, Institutional Development and Communications Northwest Regional Educational Laboratory Position:

Address: 101 S.W. Main St., Suite 500 Tel. No.: (503) 275-9517

Portland, OR Zip Code: 97204 Date: 11/15/00

III. DOCUMENT AVAILABILITY INFORMATION (Non-ERIC Source)

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents which cannot be made available through EDRS.)

Publisher/Distributor: _____

Address: _____

Price Per Copy: _____ Quantity Price: _____

IV. REFERRAL TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER

If the right to grant reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

_____

_____